

本周周报(11.5-11.11):

解聪

本周工作:

淘宝交易

本周对淘宝交易可视化的系统添加了两个视图，并完善了整个系统的分析流程。

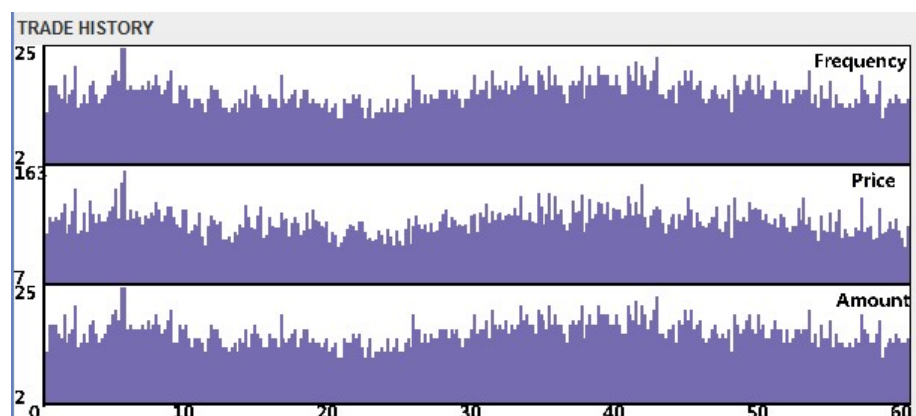
类目信息视图

类目视图显示了图例，类目的名称以及类目的相关信息。我们选取了这一时间段内交易量排名前 15 的类目予以显示，同时显示了这些类目对应的全体交易金额与交易数量。当用户在 STL 视图中点击到类目的某一行时，系统自动过滤出这一类目下的交易，可视化视图中会集中显示该类目相关的信息。对相同类目的再次点击会取消可视化视图中的类目过滤。

Category		
Category	Amount	Price
居家日用/婚...	200203	1120.26
网游装备/游...	6711	2994.89
书籍/杂志/...	1059	6146.613
女装/女士精品	827	61539.277
玩具/模型/...	504	8184.4404
女士内衣/男...	360	32062.34
移动/联通/...	299	22575.379
电子/电工	270	1328.62
童装/童鞋/...	259	15725.352
男装	231	64251.918
网络游戏点卡	168	6686.0396
3C数码配件	128	11646.51
饰品/流行首...	120	4375.1304
女鞋	106	37984.92
运动/瑜伽/...	97	6416.5
清洁/卫浴/...	85	2914.1802

用户历史数据视图

当用户可以在可视化视图中点选某一个音符时，该视图显示了用户在音符视图中点选的 ID 在最近一小时内的交易分布情况。视图包括该 ID 的频率，交易金额与交易包含商品数的分布。



另外，本周对一些交互进行了修改与完善。比如，添加了调解画布宽度的交互，调整画布中音符显示的异常程度的置信水平等。

在与淘宝的分析师的交流过程中，了解了进一步需求，并修改了系统的部分功能。目前基本可以完成两种异常的分析流程：频率异常的情况与特定类目下的价格与

数量的异常。

本周重点放在了技术报告的书写中。本周还查找了一些用户行为分析的论文，寻找是否可以有参考的地方。总的来说，现有的方法大多针对用户的历史数据做预测分析，并没有结合时间维度实时数据。自然也无法做到实时异常数据的实时监控。很多工作从用户的分析入手，类似于淘宝标签数据的分析，而从时序交易数据入手的不多。

目前现有的系统中存在视觉设计的问题外，数据分析的部分也有待改进。另外，现在只用了一小时的数据，有可能需要使用一天或者一周的数据以验证方法的准确性。

下周工作：

淘宝交易数据可视化的文档书写。

系统可视化编码的进一步完善，使得编码能够更容易理解。

将数据由一小时扩展为一天的范围，并实现层次式的可视化。